Fetal Exposure to Cigarette Smoking and Adult and Old Age Mortality: Examining the Effects of the Introduction of State Level Cigarette Taxation 1921-1940 Using Linked Full-Count Census and Mortality Records

Jonas Helgertz[1,2] and John Robert Warren[1]

[1] Minnesota Population Center, University of Minnesota

[2] Centre for Economic Demography and Department of Economic History, Lund University, Sweden

Introduction

This paper examines the intergenerational effects of smoking, more specifically how exposure to cigarette smoking while in utero influences an individual's adult and old age health, through mortality risk. A large literature has examined the association between smoking during pregnancy and in infancy or childhood, demonstrating negative effects on various aspects of the fetal growth process, where children of smoking mothers experience a significantly higher risk of being born preterm or small for gestational age. In childhood, exposure to cigarette smoking while in utero has been demonstrated to be linked to obesity and associated problems, as well as high blood pressure and a range of respiratory outcomes (Russell et al 1966, Banderali et al 2015, Jaakkola & Gissler 2004, Von Kries et al 2002). Linking exposure to cigarette smoking during gestation and its consequences for later life health emerge as considerably less studied (see, however, Bengtsson and Nilsson 2018). This paper aims to contribute to this by exploiting plausibly exogenous variation in exposure to smoking during pregnancy induced through the introduction of state level cigarette taxation in the United States, investigating its effect on adult and old age mortality. Taxation was introduced gradually across U.S. states from 1921, and this paper will focus on the cohorts born 1919-1940. Despite the time period examined being characterized by increasing smoking prevalence, the expectation is that children who were in utero at the time of the introduction of state level cigarette taxation would have enjoyed a (temporary) reduction in exposure to smoking, both directly through the mother and indirectly, through the mother's second hand smoking. The paper will rely on full-count Census data from 1930 and 1940, linked using state-of-the-art methods of probabilistic record linkage to the Social Security Death Master File (SSDMF) and to the NUMIDENT, from which death records are obtained.

The health consequences of using tobacco products are well-known today, with the CDC reporting that nearly one in five deaths is die to cigarette smoking (CDC 2018). Decades of empirical research has shown smoking to substantially increases the risk of several different types of cancers (e.g. Peto et al 2000, Gaudet et al 2013, Nomura et al 2012) and diseases that either shorten life expectancy or at the very least have a significant negative influence on quality of life, such as chronic obstructive pulmonary disease (Postma et al 2015) and asthma (Pietinalho et al 2009). Health fears concerning the use of cigarettes and tobacco more broadly have existed since long before they were scientifically proven; words of caution were offered by medical professionals and various prominent figures as early as the 1800s. Due to the long latency period between exposure and manifestation of illness, it was not, however, until the 1960's that sufficient medical evidence had accumulated to allow the Surgeon General to conclusively state that smoking represents a health hazard, through being causally linked to lung cancer (Brandt 2007).

At the time of the invention of the mass produced cigarette, in the 1890s, tobacco use had already been a phenomenon in America since centuries. While never without its opponents, on either moral or health

grounds, other uses of tobacco – as chew, snuff or in a pipe or cigar – remained dominant until into the 20th century. Through advances in the mass manufacturing of cigarettes, combined with effective, aggressive and targeted advertising, especially towards the young and women, as well as lobbying, the use of cigarettes became more and more common during the early decades of the 20th century. As cigarette sales nationwide continued to increase, with an adult in 1920 on average smoking 500 cigarettes a year, measures to prohibit or at the very least restrict sales were implemented in several states during the first decades of the century. While restricting the sales of cigarettes only to individuals above a certain age were commonplace, no less than fifteen states at some point prior to 1920 enacted a complete prohibition against the sale of cigarettes. Evidence, however, suggest that the enforcement of said prohibition was lax in the face of massive public demand. The definite break-through of the cigarette in terms of being publically accepted came during World War I, supported by organizations as the Red Cross and the YMCA supplying American soldiers in France with over 400 million cigarettes a month (Brandt 2007).

Well before the cigarette became a popular product, federal taxation was introduced in 1864. During its first year in effect, tax revenue from the sale of cigarettes represented 1‰ of the total revenue from tobacco sales. In 1920, the corresponding figure had grown to 51 percent, and at the end of the period examined, in 1940, it had grown further to 88 percent, vividly illustrating how dominant the product had become. In 1921, Iowa became the first state to implement a state tax on cigarettes, at a rate of 2 cents per pack (of 20 cigarettes). Given that the price of a pack of premium brand cigarettes at the time amounted to around 15 cents, if the tax increase was fully levied on the consumer, it would have represented a nontrivial increase in the price of cigarettes. Combined, the state and federal taxes represented approximately 50 percent of the consumer price, something that appears to have been the case until the late 1960s. Until 1940, 25 other states followed by also introducing a state tax on cigarettes. In addition, since the federal tax rate remained constant from 1919 and until 1940, the introduction of state taxes should have been the only exogenous "shock" that would have affected the price of tobacco at this level of geographic aggregation. Using data from 1960 indicate that states varied somewhat in terms of the share of the total tax that was due to the state tax, with most states, however, opting for a state tax that was between 35-40 percent of the federal tax (Orzechowski & Walker 2014).

During the decades examined in this study, cigarette smoking became pervasive in the United States. Growing sales were both the result of increased consumption as well as of the recruitment of new smokers. As already mentioned, the cigarette companies particularly targeted young women, with no previous experience of consuming tobacco. Indeed, the cigarette manufacturers successfully packaged smoking as associated with independence, attractiveness and sociability (Brandt 2007). Consequently, given the distinctive image associated with smoking as well as the vast majority of women having no prior experience from tobacco, it does not appear very likely that women who found cigarettes too expensive as a result of a price increase would choose to switch to other forms of tobacco, such as chew tobacco or cigars.

Previous research

A vast literature has mapped the importance of circumstances experienced during the fetal stage for health and other outcomes during the life-course. While a large part of the literature is associated with nontrivial concerns as far as the ability to credibly make causal claims is concerned due to the outcome and the exposure of interest being correlated with some unobserved characteristic, studies investigating the consequences of fetal exposure to exogenous environmental shocks such as the 1918 influenza pandemic (e.g. Almond 2006) or the 1944 Dutch hunger winter (e.g. Lumey et al 2009) has helped the scientific community to reach a consensus regarding the importance of the fetal stage for postnatal life. As previously

outlined, studies have demonstrated exposure to maternal (or paternal) smoking while in utero to be associated not only with birth outcomes, including low birth weight, etc, etc, but also with outcomes throughout childhood and into early adulthood. While it has also been shown that several of aforementioned outcomes are linked to adulthood and old age mortality, to our knowledge, no study has directly examined the link between fetal exposure to cigarette smoking and old age mortality.

A key issue for the purpose of this paper is i) that the introduction of taxes represents are a cost that is (at least partially) borne by the consumer, and that ii) consumers react to price increases by reducing their consumption. While no study has empirically studied the consequences of the introduction of the cigarette taxes examined in this paper, more recent taxation changes have been the subject of study, with no consensus reached regarding whether the tax burden is entirely shifted over to the consumer. Despite the inconsistencies in the results, the consensus is that the tax burden is at least partly shifted over to the consumers. Using U.S. data from the time period 1955-1982, Ashenfelter and Sullivan (1987) uses the combined state and federal tax rate finding that the price increase did not amount to the amount of the tax increase. Their results are at odds with Barzel (1976) who finds that a one cent increase in the state tax translates to a 1.07 cent increase in the price of cigarettes. This result is confirmed by Hanson and Sullivan (2009) and Sullivan and Dutkowski (2012), finding that tax increases are over-shifted to the consumers. The former study is furthermore of relevance here, as it shows variation in the extent to which the burden is shifted over to the consumer by the distance to a neighboring state, depending on the price in the adjoining state.

Turning to the effect of price increase on cigarette consumption, to our knowledge, no study has examined this relationship during the context and period of relevance for this paper. Furthermore, due to the demand being driven also by addiction, cigarettes represent a particular commodity which may be governed by different demand mechanisms than a normal good. Existing studies have been conducted on a more recent time period, however, paint a rather consistent picture suggesting that consumption diminishes as a result of increases in the price. Licari and Meier (1997) examine U.S. data for the period 1951-1994 and find that a one percent real increase in state taxes of cigarettes reduces the consumption with .81 packs per person. These results are confirmed by Chaloupka et al (2002), also showing, using U.S. data, that those most affected by a price increase are younger and/or low income/education individuals (see also Powell et al 2005).

Data

The key independent variable for the study is represented by time series data on the date of the introduction of state level taxation on cigarettes, starting with Iowa in 1921 (Orzechowski & Walker 2014). We focus on the time period until 1940, during which 25 states followed and implemented similar taxation laws. While the taxation data is a valuable source of information, it is not without its (possible) caveats. The source does not provide time series data on the amount of tax that was levied or how (if) it changed during the time period examined. Time series data on the amount of state tax is, however, available from 1955, from which it can be observed that the tax was 1955 is within +/- one cent of the tax when first introduced. Thus, there should only have been negligible changes across time in the tax rate.

The study sample is extracted from the U.S. full count Censuses of 1930 and 1940, where our study population is represented by individuals born between 1919 and 1940[1]. Exposure to the introduction of state taxes while in utero is defined through the use of information on state of birth and the individual's date of

[1] From 1930, we extract individuals born 1919-1929, and from 1940, we extract those born 1930-1940.

birth provided by the NUMIDENT/SSDMF (explained further below). Thus, this allows for the precise definition of whether an individual was in utero at the time of the implementation of the tax. From the Census, we will also exploit information on the individual's socioeconomic background, in order to investigate whether responses to the price increase differ by the parents' socioeconomic status. Lastly, for cohorts born 1919-1929 (1930-1940), information on place of residence in the 1930 (1940) census will be used to approximate the distance to the nearest neighboring state.

The outcome of interest is represented by the individual's age at death, provided by the NUMIDENT/SSDMF databases. The NUMIDENT, containing 50,000,000 unique death records, represent the main source of death data. While containing fewer deaths than the SSDMF (~65,000,000), it also contains individual level data from Social Security claims unrelated to the individual's death, which provides us with additional information. Since both sources contain Social Security Numbers, we will complement the NUMIDENT with deaths from the SSDMF, should they not be recorded in the first source.

Methods

Record linkage

The absence of a common identifier across the datasets of individual level data that will be used in the paper means that individuals selected from the 1930/1940 census need to be linked to their death record in the NUMIDENT/SSDMF using methods of probabilistic record linkage. This represents a field that is developing rapidly within the social sciences, where the fundamental underlying idea when linking individuals across large scale databases, such as full-count censuses, is one where the researcher "trains" a machine (i.e. the computer) to recognize patterns in the data that are consistent with two records being uniquely similar to the extent that they represent the same person. In theory, linking individuals with unique names represents less of a challenge; a Frank Oberhauser should be reasonably easy for the computer to find across records, whereas a John Smith represents an individual that appears to be virtually impossible to link, due to the inability to distinguish between the many possible matches. In order for the linking algorithm to be calibrated, it requires accurately coded training data. More specifically, the training data is a representative subsample of the greater population that the researcher aims to link, where potential links have been manually assessed. Using a train-test-split procedure, where the machine learning algorithm is calibrated on one part of the data and tested on the other (where the true matches are known) its performance can be assessed. Here, the researcher will put a particular emphasis on training the algorithm to minimize the number of false positives (i.e. declared matches that, in reality, are non-matches) while at the same time maximizing the number of true positives (i.e. declared matches that are actual matches).

This paper will implement a linking strategy that partly redefines the role of training data. The primary logic of this being that the evaluation of the performance of any algorithm only is as good as the training data on which it was based. Consequently, an algorithm that performs very well based on training data which contains incorrectly declared matches will provide the researcher with a misleading assessment regarding how well it is actually performing. This paper uses a novel way of performing record linkage which aims to enhance the quality of the training data and, hence, also of the resulting linked data. The point of departure of the approach is to rely on more complete family level information when declaring links rather than only use information on the index individual. Furthermore, this information is taken into account not only when creating the training data, but also explicitly when linking the full databases, i.e. when formally modelling the relationship between the sources.

One example of how this differs from methods used by Ferrie (1996) and Feigenbaum (2016) is illustrated below. Methods currently used links a record from source A to all possible records (conditional on blocking variables, i.e. sex, state of birth and year of birth/age) in source B. An Arthur Thomas from source A, born

1922 in Arizona would typically be linked to all males born +/-3 years in Arizona in source B. While not a very unusual name, let's assume that the linking procedure yields four potential matches with that name (also allowing for trivial spelling errors). According to both the Ferrie and Feigenbaum approach, it would be impossible to distinguish between the matches. The novelty introduced in this paper is the addition of information on the individual's parents, available in both the Censuses and in the NUMIDENT. It is easy to see that this information allows us to distinguish between the alternatives, information that also will be fed into the linking algorithm through formally modeling also parental information.

1900								1910							
ID last name	ID first name	Mother first	Mother year of birth	Mother place of birth	Father first	Father year of birth	Father place of birth	ID last name	ID first name	Mother first	Mother year of birth	Mother place of birth	Father first	Father year of birth	Father place of birth
thomas	arthur	mary a	1872	41000	benjamin h	1870	900	thomas	arthur	nellie g	1872	900	andrew d	1868	41000
thomas	arthur	mary a	1872	41000	benjamin h	1855	41000	thomas	arthur	katherine m	1855	41000	edward j	1868	41000
thomas	arthur	mary a	1872	41000	benjamin h	1867	41000	thomas	arthur	mary a	1873	41000	benjiman h	1868	41000
thomas	arthur	mary a	1872	41000	benjamin h		900	thomas	arthur	julia	900	900		1868	41000
thomas	arthur	mary a	1872	41000	benjamin h	1867	3600	thomson	arthur	mary j	1867	15000	thos k	1868	41000
thomas	arthur	mary a	1872	41000	benjamin h	1861	40000	thomsen	arthur	marie	1864	40000	christian	1868	41000
thomas	arthur	mary a	1872	41000	benjamin h	1862	3600	thompson	arthur	agusta	1866	900	edward	1868	41000
thomas	arthur	mary a	1872	41000	benjamin h	1863	40000	thompson	arthur	anna	1871	45300	soren	1868	41000

The example outlined above is not the only of how using this information improves the quality of the training data, and as a consequence, the usefulness of the model fit statistics and thus also the overall linking algorithm. Our data suggests that the use only of individual level information substantially overestimates (underestimates) the number of false negatives (false positives) in the training data, with nontrivial consequences for the overall linking performance.

Empirical strategy

The baseline analysis is performed using differences-in-differences, following equation (1), below.

$$Y_{ist} = \gamma_s + \mu_t + \beta D_{st} + \rho X_i + \varepsilon_{ist}$$

Y_{ist} is the outcome, referring to the age at death of individual *i*, born in state *s* at time *t*. γ_s represents state of birth fixed effects, whereas μ_t represents birth cohort fixed effects. The parameter of interest, βD_{st} captures the effect of the introduction of cigarette taxation, which takes the value 1 if the tax has been implemented for in-utero cohort *t* in state *s*, 0 otherwise. The vector X includes a range of control variables, where a particular focus will be directed to the influence of the individual's parents socioeconomic status, a measure of the Euclidean distance to the neighboring state, as well as a measure of between-state differences in the price level of cigarettes. Lastly, ε_{ist} represents an individual specific error term.

In order to further investigate the nature of the effect, its intensity will be allowed to differ according to the individual's socioeconomic background as well as by distance to the state border. As previously outlined, the harmful health effects of cigarette smoking were unknown, why the response to a price increase due to the introduction of cigarette tax is presumed to be a function of the individual's degree of addiction as well as their ability to continue financing the habit. The time period examined, especially since the early 1930s, was characterized by widespread smoking, across gender, socioeconomic status or region. As regards the latter, the price increase should primarily have been felt among low SES individuals, thus being more of an incentive to reduce or stop consuming cigarettes than among high SES individuals. The geographical dimension of the effect of cigarette taxation on consumption refers to the expectation that individuals in affected states but who live near the border to an unaffected state should be more likely to have access to acquiring lower priced cigarettes. The presence of bootlegging, i.e. the illegal and organized smuggling of cigarettes across the border, has been empirically established, as has the existence of price differences on

legal cigarettes according to the distance to areas characterized by lower prices (Thursby and Thursby 2000).

References

Almond, D. 2006. "Is the 1918 Influenza Pandemic Over? Long-term effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." Journal of Political Economy, 114:672-712.

Russell, C.S., Taylow, R., & Maddison, R.N., (1966): Some Effects of Smoking in Pregnancy. J Obstet. Gynaec. Brit. Cwlth., 73, 742-746

Banderali, G., Martelli, A., Landi, M., Moretti, F., Betti, F., Radaelli, G., Lassandro, C., Verduci, E., (2015): Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. J Transl Med, 13: 327

Jaakkola, J.J.K., Gissler, M., (2004): Maternal Smoking in Pregnancy, Fetal Development, and Childhood Asthma. American Journal of Public Health, 94(1).

Von Kries, R., Toschke, A.M., Koletzko, B., Slikker, W., (2002): Maternal Smoking during Pregnancy and Childhood Obesity, American Journal of Epidemiology, 156(10).

Peto, R., Darby, S., Deo, H., Silcocks, P., Whitley, E., Doll, R., (2000): Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. BMJ, 321.

Gaudet, M.M., Gapstur, S.M., Sun, J., Diver, R., Hannan, L.M., Thun, M.J. (2013): Active Smoking and Breast Cancer Rink: Original Cohort Data and Meta-Analysis. J Natl Cancer Inst, 105.

Nomura, A.M.Y., Wilkens, L.R., Henderson, B.E., Epplein, M., Kolonel, L.N., (2012): The association of cigarette smoking with gastric cancer: the multiethnic cohort study. Cancer Causes Control, 23(1).

Postma, D.S., Bush, A., van den Berge, M., (2015): Risk factors and early origins of chronic obstructive pulmonary disease. Lancet, 385.

Pietinalho, A., Pelkonen, A., Rytilä, P., (2009): Linkage between smoking and asthma. Allergy, 64.

Brandt, A.M., (2007): The Cigarette Centrury. Basic Books.

Orzechowski, Walker (2014): The Tax Burden on Tobacco. Historical Compilation 2014.

Lumey, L.H., Stein, A.D., Kahn, H.S., Romijn, J.A., (2009): Lipid profiles in middle-aged men and women after famine exposure during gestation: the Dutch Hunger Winter Families Study. Am J Clin Nutr, 89.

Powell, L.M., Tauras, J.A., Ross, H., (2005): The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. Journal of Health Economics, 24.

Barzel, Y., (1976): An Alternative Approach to the Analysis of Taxation, 84(6).

Sullivan, R.S., & Dutkowsky, D.H., (2012): The Effect of Cigarette Taxation on Prices: An Empirical Analysis Using Local-Level Data. Public Finance Review, 40(6).

Hanson, A.R., Sullivan, R.S., (2009): The Incidence of Tobacco Taxation: Evidence from Geographic Micro-Level Data. National Tax Journal 62.

Licari, M.J., Meier, K.J, (1997): Regulatory Policy When Behavior is Addictive: Smoking, Cigarette Taxes and Bootlegging. Political Research Quarterly, 50(1).

Chaloupka, F.J., Cummings, K.M., Morley, C.P., Horan, J.K., (2002): Tax, price and cigarette smoking: evidence from the tobacco documents and implications for tobacco company marketing strategies. Tobacco Control, 11(S1).

Ferrie J.P., (1996): A new sample of males linked from the Public-Use-Microdata-Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census manuscript schedules. Historical Methods. 29.

Thursby, J.G., Thursby, M.C., (2000): Interstate Cigarette Bootlegging: Extent, Revenue Losses and Effects of Federal Intervention. National Tax Journal, 53(1).